



Europäisches Patentamt  
European Patent Office  
Office européen des brevets



⑪ Publication number: **0 554 492 A1**

⑫

## EUROPEAN PATENT APPLICATION

⑳ Application number: **92102049.1**

⑤① Int. Cl.<sup>5</sup>: **G06F 3/00**

㉔ Date of filing: **07.02.92**

Amended claims in accordance with Rule 86 (2) EPC.

④③ Date of publication of application:  
**11.08.93 Bulletin 93/32**

⑧④ Designated Contracting States:  
**DE FR GB**

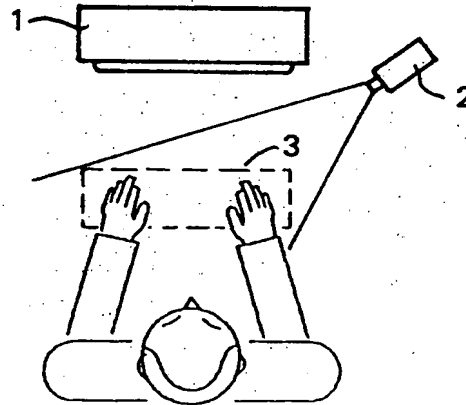
⑦① Applicant: **International Business Machines Corporation**  
**Old Orchard Road**  
**Armonk, N.Y. 10504(US)**

⑦② Inventor: **Korth, Hans-E., Dipl.-Phys.**  
**Sandbergerstrasse 34**  
**W-7000 Stuttgart(DE)**

⑦④ Representative: **Klocke, Peter, Dipl.-Ing.**  
**IBM Deutschland Informationssysteme GmbH Patentwesen und Urheberrecht**  
**Pascalstrasse 100**  
**W-7000 Stuttgart 80 (DE)**

⑤④ **Method and device for optical input of commands or data.**

⑤⑦ A method of user command or data acquisition in a processing system is described that allows to analyze human hand and finger motions. The activities are then interpreted as operations on a physically non-existent input device. The resulting optical user/computer interface complements the visual screen-to-user path with a hand-to-image data acquisition link from the user to the processor unit. It allows to emulate different types of manually operated input devices (mouse, keyboard, etc.). Mechanical input units can be replaced by virtual devices, optimized for the current application and for the user's physiology. The speed, simplicity and unambiguity of manual data input will be maintained. Sensing of finger motions can generate feedback to reduce the hazard of repetitive strain injury (RSI) syndrome.



**FIG. 1**

The present invention lies in the field of interfaces for direct data or command input by the human user. More specifically, the invention concerns devices which are operated manually and react on finger tip pressure or other hand induced motions.

To enter data e.g. into a computer, the keyboard is the most widely used device. It allows fast, unambiguous input of digital data. Alternative input devices may be used for special, limited applications: Voice recognition is inherently fast, but requires a very large processing overhead. Ambiguity is inherent to the voice-to-text conversion. Automated hand-writing analysis is difficult, too: Input of easily readable single characters is slow and tedious. To read individual writing can be a complex and time consuming task.

Pointing devices (mouse, pen) have found widespread application in conjunction with graphical windowing environments to control the processing and to enter binary data. With this arrangement a typical computing session requires frequent switching between pointer and keyboard.

In CAD applications typically a mouse is used for editing graphics. Other devices, e.g. joysticks or trackballs can be added to the system to allow convenient entering of viewing commands, such as rotation of a virtual 3D-object for generating a drawing of another perspective.

A specific optical input device for use in a graphic station is described in IBM TDB Vol. 32, No. 3B, August 1989, p. 92. This device consists of a hand held cube whose faces are painted specifically to provide cues for surface orientation. Actual position and orientation of the cube is detected by a TV-sensor. The recorded picture of the cube is digitally processed and commands are extracted to change the graphic presentation of a virtual 3D-object in correspondance to the cube's orientation. There are remaining problems inherent in this technique as due to the fact that the cube is held by hand the orientation cues are partially obscured.

With ongoing trend to miniaturisation advanced computers are characterized by reduced dimensions. While the size of the electronics is shrinking, the dimensions of the user interface, e.g. keyboard or mouse, cannot be reduced too much without a loss of functionality.

The present invention is intended to remedy the drawbacks inherent with the prior art user input devices.

It is therefore object of the present invention to provide a new and advanced interface for data or command input in data or command processing systems which avoids problems inherent with mechanical components and allows to emulate various types of manually handled input devices.

This is achieved by the present invention as claimed, disclosing an optical user interface which in general is an arrangement using an image acquisition system to monitor the hand and finger motions and gestures of a human user. These activities are interpreted as operations on a physically non-existent computer keyboard or other input device. The image acquisition system comprises a TV-sensor and means for digitizing and processing signals from said TV-sensor. Hand, fingers and finger-tips of the user are observed by the TV-sensor and active motions are detected and analyzed by said means for processing signals and interpreted as respective input data or commands.

The required hardware for the inventive optical interface is just a (e.g. a CCD-array) video sensor with 'frame grabbing' circuitry. A 100x100 pixel array should be the minimum for the required image processing. An anamorphic imaging lens or a sensor chip with stretched pixels may be used to match the sensor with the aspect ratio of the keyboard. Autofocus or fixfocus will simplify the setup procedure. Standard sensors for camcorders or electrophotographic sensors easily fulfill these requirements.

Black and white imaging will be sufficient, if the desktop can be discriminated by a grey-level that differs significantly from the shade of human skin. If no assumptions can be made about the background (e.g. desk surface), some low-level image processing is required to determine the contours of the operator's hands. With the a priori knowledge that the background image remains constant (except for noise and shading) and the knowledge about the hand as a contiguous entity, only a moderate amount of time-critical processing will be necessary.

It should be noted that due to the lower part count the costs for the camera and the frame-grabbing hardware can ultimately be lower than the cost for a good stand alone keyboard with driving circuitry. A further cost reduction arises from the fact, that some defective pixels (e.g. on the CCD-chip) can be tolerated.

A TV camera with 50 (half-) frames per second will be fast enough to monitor typing speeds of some 10 keys per second. While general image processing tasks require a large amount of computing power, keystroke detection can be done quite economically, since there is a lot of a priori knowledge involved.

If the camera is directed e.g. from a monitor or a computer housing towards the operator (Figure 1), the lower part of the image is filled by the working desk. Objects moving over the desk can be identified as hands quite easily by their size, shape and motion. When a hand has been identified, its displacement within subsequent image

frames will be rather small. Therefore, only a moderate number of pixels must be analyzed to keep track of the motion and to verify the integrity of the picture scene.

Given the shape of a hand over an imaginary keyboard, a simple model allows to attribute the finger-tip locations to the (up to five) protrusions on the lower side of the hand contour. The motions of the fingers relative to the hand are small too. They can be monitored efficiently by the evaluation of small image regions. The steadiness of the motions allows to extrapolate the locations of the search regions.

Obviously, in addition to the finger-tip motions the motions and the shape of the hands can be used to send gestures with additional information to the optical interface (e.g., begin/end of session, keyboard redefinition, camera mode, or other commands). Thus the invention realizes an advanced contactless data/command input device, capable to replace manually operated means in data processing systems or electronic compounds in general.

The optical user interface can be used for a multitude of tasks. They can be exchanged instantly and can be switched with the current application.

**Alphanumeric keyboards:** A striking advantage is offered by a virtual keyboard for 'sub-notebook' sized computers. A full-size keyboard is available with the compact device.

**Keyboards for large character sets:** The visualization of the actual key function allows the handling of keyboards with multiple shift keys and subgroup display for large character sets (e.g. Kanji).

**Calculator panels:** The layout can be optimized, e.g. for input speed, simplicity or an appropriate set of function keys.

**Arbitrary analog and digital control panels:** Different control panels, e.g. trackball, joystick can be emulated for the actual task.

**Keyboards for musical instruments:** Equipped with a virtual piano keyboard, a PC becomes a complete system for music production. External keyboards, synthesizers/sequencers will be obsolete. Impact dynamic sensitivity is inherent to the sensor algorithm. With camera and piano keyboard a PC will have a 'Multi-media' interface allowing input, output, processing and storage of pictures and tunes.

**Personal signature identification:** A positive ID can be obtained from signature dynamics.

**Alternative camera input:** The image acquisition capability of the system can be used additionally for other purposes, e.g. document scanning, FAX, picture-phone.

For a better understanding of the present invention, together with further objects and advan-

tages, as a preferred embodiment of the invention the use of the optical interface as a virtual keyboard is described in the following with reference to the accompanying drawings, in which:

Fig. 1 shows the arrangement of the image acquisition system to detect the hand and finger motions of a human user.

Fig. 2 is a schematic plot of the camera image together with an enhanced area segmentation of an observed hand.

Fig. 3 demonstrates the steps of hand image processing.

The task of a keyboard is to provide information about the available keys and their location, to sense the depression of keys by the operator and to produce some feedback when a key has been touched.

The keyboard needs not to exist physically. In Fig. 1 a virtual keyboard 3 serves as data or command input device. From the computer side of the interface there must just be a way to detect the fingertip motions of the keyboard operator. Conventionally, this is done with electrical contacts below the individual keys. However, other approaches will work likewise: If the hand and finger movements on a virtual keyboard (e.g., a table-top) are monitored by an optical system, appropriate image processing can detect the user's intent.

Fig. 2 shows the area viewed by camera 2. In the observed region 4 the user's hands are detected. The enlarged section 5 of Fig. 2 represents an area of the camera's image which is chosen for further image processing.

A 3-level algorithm minimizes the execution time (Fig. 3). The 1st level algorithm scans contrast values along a line 6 to detect presence of a hand. The 2nd level routine follows the hand contour 7. The 3rd level algorithm finds the finger tip locations (81, 82, 83, 84, 85).

Key information for a virtual optical keyboard 3 (VOK) can be displayed in a window on the monitor screen 1. Schematic representations of the hand images can be displayed over a key template. The simultaneous display of the keyboard with the application will be advantageous for ergonomic reasons, since there is no need to switch the view between screen and keyboard.

The keyboard window may require some 10% to 20% of the display area. However, the window may appear only during the input of data, so that other simultaneous applications will not be affected. Checking the input cursor location, the keyboard window may position itself on the screen opposite to the text input area.

A finger-tip hitting a virtual key does a rapid movement downward (over a path of some 10mm), stops abruptly when the desk surface is hit and moves back upward after a brief relaxation. The

location where the hitting finger-tip stops can be attributed to a key. Ambiguities due to simultaneously moving fingers can be resolved evaluating the different motion characteristics (the hitting finger moves faster over a longer path and stops more abruptly) and by a hand model that indicates unlikely relative finger positions.

In practice, the finger tip position can be determined for every frame (or half-frame), i.e. about every 40 msec. A sequence of five frames allows to monitor the tip motion for a 160 msec period. The tip location difference between two subsequent frames then gives four values for the finger speed ( $V_4 \dots V_1$ ). To identify a hit, a sequence of speed values has to be found that corresponds to the description above. To solve this problem with only four speed measurements some assumptions shall be made:

- Before the impact the finger tip acceleration will be relatively constant. This seems reasonable, because the acceleration depends only on the muscle tension. In this case, the speed ( $V_x$ ) near the impact point can be calculated by extrapolation from the two pre-hit speed values. The extrapolated value can then be compared with a threshold value.
- The hit onto a virtual key implies, that the pre-hit speed values must be above a minimum value.
- On impact the speed decreases rapidly. Therefore, the speed measurement during the impact ( $V_2$ ) does not contain helpful information. It may be used, however, to check the validity of the measurement sequence (absence of noise).
- After the impact the finger tip speed decreases rapidly. The deceleration depends on elastic finger tip deformation. The post-hit speed must be within a range determined by the minimum deceleration and the maximum of the following reacceleration.

With these assumptions a set of four simple conditions can be written that must be fulfilled for a hit:

1.  $V_x > V_{min}$  ( $V_{min}$  = minimum impact speed)
2.  $V_4 > V_p$  ( $V_p$  = minimum pre-hit speed)
3.  $V_2 > V_1$  no oscillation
4.  $V_1 < V_t$  ( $V_t$  = maximum post-hit speed)

When a key hit has been detected, the corresponding key has to be found. This is not a trivial task, as the finger is moving. If the finger is guided visually, the key selection is done before the hit motion starts, i.e. some 100 msec before the hit detection. For fast typing under visual control, the 'search and hit' impulse of the human operator is generated from the visible distance between finger and the key to be hit. In this case, on the virtual key board template that the actual impact point

should be selected. In practice, a steady transition between these extremes must be implemented. Knowledge of the fingertip path allows to refine the prehit speed and acceleration criterion for the hit detection algorithm.

A self-centering routine is required to take over the tactile key center feedback provided by a physical keyboard. The human operator wants to hit the key enter. This allows an algorithm to reposition the finger image accordingly. To avoid an undesired drift of the hand image, a sort of 'stiffness' has to be implemented.

To monitor the finger tip dynamics, information about the vertical motion history of every finger tip is buffered. An uninterrupted sequence of five tip measurements is required to make a finger eligible for the hit determination algorithm. However, during the monitoring, the number of positively identified finger tips may change (e.g. due to poor illumination, fingers outside the video frame area or overlapping fingers). In this case the finger number assignment may change and the finger history will be corrupted. To avoid a dead-time of five frames for a new build-up of the history buffers, a reassignment algorithm is required: When the finger count is increased the new fingers must be identified. Pointers must be generated that allow the correct access to the history buffers. In case of a decreasing finger count, the pointers for the remaining fingers must be updated. With the virtual keyboard routines all fingers (except the newcomers) are tracked permanently. Even in case of noise induced finger-count fluctuations the fingers can be monitored for the hit signature.

A set of simple intuitive 'gestures' may be defined to control the operation of a virtual keyboard, e.g.:

- Start/End: The image processing algorithm checks for the presence of a hand, scanning the 'wrist line'. This is done e.g. every 320 msec or when a hand contour detection error occurs. If a hand is detected the keyboard template window will be opened on the display screen. If the hands are removed the window will be closed.
- Type-start position: The placement of a hand on the desk (i.e. a simultaneous five finger hit) can be interpreted as a command to place the virtual keyboard so that, for example, the left hand is atop the A-S-D-F-space keys.
- Undo: A rapid upward finger motion after a key hit will be a simple signal to 'undo' the hit. Highly intuitive will be the monitoring of the thumbs. Unconsciously, the thumbs are driven upward by the surprised reaction on a typing error. The undo gesture simplifies the keyboard action: The suspension of the cur-

rent work in order to find the back-space or the delete key can be avoided.

- Mode switch: Various gestures of one or both hands can be envisioned to switch between different keyboard modes (e.g. typewriter, mouse, calculator, finger pen/paint, host session, piano). The immediate switching between typewriter and pointing device allows a faster cursor placement.

Instead at the desk plane, the virtual keyboard can be defined relative to the finger-tip positions. This means, that the key for each finger will be assigned before the key is actually hit. A short, rapid downward movement then selects the input key. Pre-hit key assignment offers some unique features for operator feedback:

- The key assignment can be displayed on-screen (e.g. as color coding on a keyboard layout window).
- Electronically supported and stabilized key assignments allows to help operators with handicaps of the hand and finger motion.
- Enlarged or acoustical feedback can support visually impaired operators.
- The actual key function (shift-state etc.) can be indicated.
- A 'lift-off' finger motion can be detected. This will be a fast and highly intuitive way to undo typing errors.
- The real finger-tip location relative to the key center can be monitored (color, click loudness and pitch).
- The computer can dynamically adjust the center of the virtual keys with respect to the actual finger and hand locations. A nearly stationary finger will be 'placed' automatically atop the virtual key. From this location an experienced typist will find the other keys without visual control.
- Operator guidance allows to issue warning or encouragement when special keys come into reach.

Pre-hit key assignment can provide for efficient operator feedback. The virtual keyboard technique may be used to generate standard size keyboards for 'sub-notebook' size computers. Other implementations will reduce the desk surface requirements for PCs, define remote and protected keyboards and are adaptable for visually and motorically impaired persons.

Image (hand shape) segmentation becomes very simple, if a color sensor can be used and the hands are moved atop of a monochrome (blue) desk template. However, black and white imaging should work equally well, if the desktop can be discriminated by a grey-level that differs significantly from the shade of human skin. If no assumptions can be made about the desk background,

some low-level image processing is required to determine the contours of the operator's hands. With the a priori knowledge, that the background image remains constant (except for noise and shading) and the knowledge about the hand as a contiguous entity, only a moderate amount of time-critical processing will be necessary.

As mentioned above, a keyboard should provide the user with the key information and feedback. Tactile feedback of a virtual keyboard is given by the contact of the finger-tip with the working plate. Acoustic feedback (clicks) can be generated easily by the processing unit.

Key information for a virtual keyboard can be displayed in a window on the monitor screen. Schematic representations of the hand images can be displayed over a key template. The simultaneous display of the keyboard with the application will be advantageous for ergonomic reasons, since there is no need to switch the view between screen and keyboard (Note: This is quite similar to today's use of a mouse pointer).

The sensing of the finger motions can be used to generate feedback information for the operator, if the finger-tip impact is unnecessarily high. This will allow to reduce the hazard of stress induced RSI (repetitive strain injury) syndrome. In any case, since there are no physical spring-loaded keys the virtual keyboard may be operated with reduced finger strain.

A virtual keyboard allows an individual optimization of the keyboard size. This is absolutely not trivial. The hand size of humans varies over a significant range. The standard keyboard size is a compromise. When the visual, tactile and motoric functions of the operator to keyboard interface are separated, these functions can be optimized for ease-of-use and as a consequence for operating speed.

In addition, a virtual keyboard can be duplicated or split easily into separate elements for the right and the left hand. This allows a highly relaxed working position (e.g., in a chair with arm rest). The implications for handicapped persons are obvious.

For operators with not so much expertise in computer usage and typing, a printed template (place-mat) with the keyboard layout will be useful. Another image evaluation procedure then helps the image processor to match the virtual keyboard with the template. The place-mat may be applied on a piece of soft tissue (e.g. polyurethane foam) to reduce the strain on the finger-tips from the impact on a rigid desk plate.

## Claims

1. Method of user command or data acquisition in a command or data processing system where-

by user induced motions are observed, said system including a TV-sensor (2) and means for digitizing and processing signals from said TV-sensor

characterized in that

hand and/or fingers of the user are observed by said TV-sensor,

shape, position and active motions are detected and

analysed by said means for processing signals,

whereby said motions are interpreted as respective input data or commands.

2. Method according to claim 1, wherein analysis of said shape and position includes localisation of fingertips.

3. Method of claim 1 or 2, wherein said processing system additionally comprises a monitor (1) and at least a part of the display-area of said monitor is used to visualize at least schematically hand or finger motions in a respective region (4) in front of said TV-sensor (2).

4. Method of claim 3, wherein said part of said display-area is activated automatically by positioning a hand in said region (4) in front of said TV-sensor (2).

5. Method of claim 3 or 4, whereby schematic representations of the observed hand, fingers or fingertips are displayed on said monitor.

6. Method of claim 5, wherein said schematic representations were placed and moved on said monitor in correspondence to the actual position of the hand, fingers or fingertips in front of said TV-sensor.

7. Method of any one of the preceding claims, wherein analysis of said active motions includes classification by comparison of detected motions with predefined and stored patterns of motions and discrimination of not matching motions so that only specific types of motions are accepted.

8. Method of claim 7, wherein said accepted specific types of motions are interpreted as commands.

9. Method of any one of claims 3 to 8, wherein said monitor (1) additionally displays icons or symbols representing data values or commands which can be selected directly.

10. Method of claim 9, wherein said icons are chosen and arranged so as to represent a keyboard.

11. Method of any one of the preceding claims, wherein analysis of said active motions includes filtering to suppress minor motions evoked e.g. by distortions such as vibrations etc.

12. Data or command input device wherein hand or finger motions of a user are detected and analysed as recited in any one of the preceding claims.

13. Data processing system comprising at least one input device as recited in claim 12.

#### Amended claims in accordance with Rule 86 (2) EPC

1. Method of user command or data acquisition in a command or data processing system whereby user induced motions are observed, said system including a TV-sensor (2) and means for digitizing and processing signals from said TV-sensor

characterized in that

hand and/or fingers of the user are observed by said TV-sensor,

whereby the placement of said TV-sensor allows an unobstructed view at said hand and/or fingers,

shape, position and active motions are simultaneously detected by said TV-sensor and

analysed by said means for processing signals,

whereby said motions are interpreted as respective input data or commands.

2. Method according to claim 1, wherein analysis of said shape and position includes localisation of fingertips.

3. Method of claim 1 or 2 wherein said active motions comprise horizontal and vertical motions and fingertip dynamics.

4. Method of any of the preceding claims, wherein said processing system additionally comprises a monitor (1) and at least a part of the display-area of said monitor is used to visualize at least schematically hand or finger motions in a respective region (4) in front of said TV-sensor (2). 5
5. Method of claim 4, wherein said part of said display-area is activated automatically by positioning a hand in said region (4) in front of said TV-sensor (2). 10
6. Method of claim 4 or 5, whereby schematic representations of the observed hand, fingers or fingertips are displayed on said monitor. 15
7. Method of claim 6, wherein said schematic representations were placed and moved on said monitor in correspondence to the actual position of the hand, fingers or fingertips in front of said TV-sensor. 20
8. Method of any one of the preceding claims, wherein analysis of said active motions includes classification by comparison of detected motions with predefined and stored patterns of motions and discrimination of not matching motions so that only specific types of motions are accepted. 25 30
9. Method of claim 8, wherein said accepted specific types of motions are interpreted as commands. 35
10. Method of any one of the preceding claims, wherein  
said active motions comprise key hit motions,  
with said key hit motions being detected by subsequent frames of said TV-sensor. 40
11. Method of claim 10, wherein detecting said key hit motions comprises prehit key assignment by observing the fingertip dynamics. 45
12. Method of any one of claims 4 to 11, wherein said monitor (1) additionally displays icons or symbols representing data values or commands which can be selected directly. 50
13. Method of claim 12, wherein said icons are chosen and arranged so as to represent a keyboard. 55
14. Method of any one of the preceding claims, wherein analysis of said active motions includes filtering to suppress minor motions evoked e.g. by distortions such as vibrations etc.
15. Data or command input device wherein hand or finger motions of a user are detected and analysed as recited in any one of the preceding claims.
16. Data processing system comprising at least one input device as recited in claim 15.

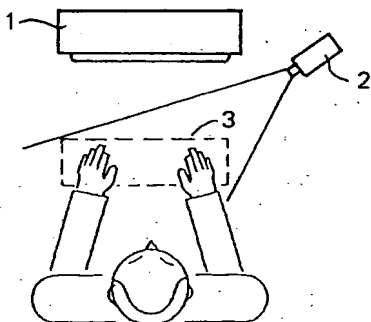


FIG. 1

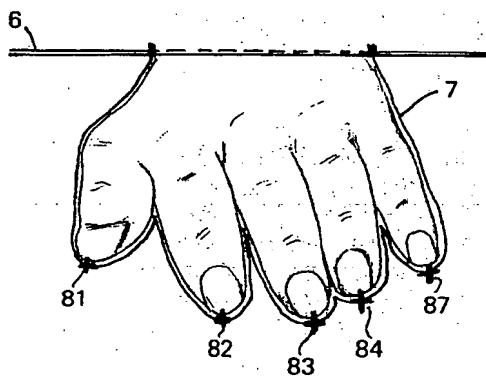


FIG. 3

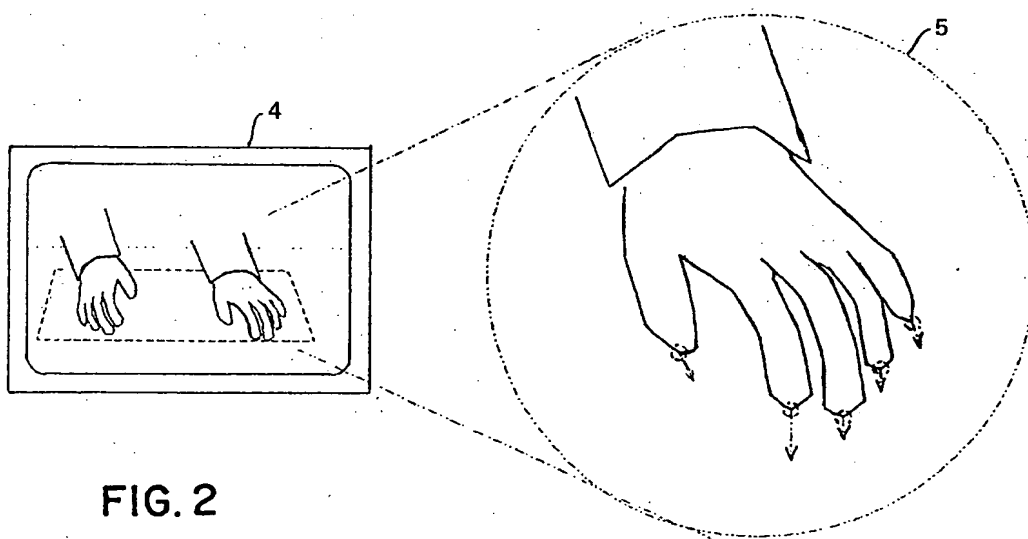


FIG. 2





European Patent  
Office

## EUROPEAN SEARCH REPORT

Application Number

EP 92 10 2049

DOCUMENTS CONSIDERED TO BE RELEVANT			
Category	Citation of document with indication, where appropriate, of relevant passages	Relevant to claim	CLASSIFICATION OF THE APPLICATION (Int. Cl.5)
X	PATENT ABSTRACTS OF JAPAN vol. 8, no. 259 (P-317) 28 November 1984 & JP-A-59 132 079 ( NIPPON DENSHIN DENWA KOSHA ) 30 July 1984 * abstract *	1,2,7,8, 11-13	G06F3/00
X	--- PATENT ABSTRACTS OF JAPAN vol. 14, no. 360 (P-1088) 12 November 1988 & JP-A-21 032 510 ( SONY CORP ) 22 May 1990 * abstract *	1-3,5-9, 12,13	
X	--- PATENT ABSTRACTS OF JAPAN vol. 12, no. 439 (P-788) 5 January 1987 & JP-A-63 167 923 ( PFU LTD ) 12 July 1988 * abstract *	1-3,5, 12,13	
A	--- US-A-4 843 568 (MYRON W KRUEGER) * page 11, line 21 - line 47; figures 2,18 * * column 25, line 36 - column 26, line 11 * -----	1,2	
			TECHNICAL FIELDS SEARCHED (Int. Cl.5)
			G06F
The present search report has been drawn up for all claims			
Place of search THE HAGUE		Date of completion of the search 25 SEPTEMBER 1992	Examiner ALONSO Y GOICOLEA L.
<b>CATEGORY OF CITED DOCUMENTS</b> X : particularly relevant if taken alone Y : particularly relevant if combined with another document of the same category A : technological background O : non-written disclosure P : intermediate document T : theory or principle underlying the invention E : earlier patent document, but published on, or after the filing date D : document cited in the application L : document cited for other reasons * : member of the same patent family, corresponding document			